



The
Patent
Office



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ



I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

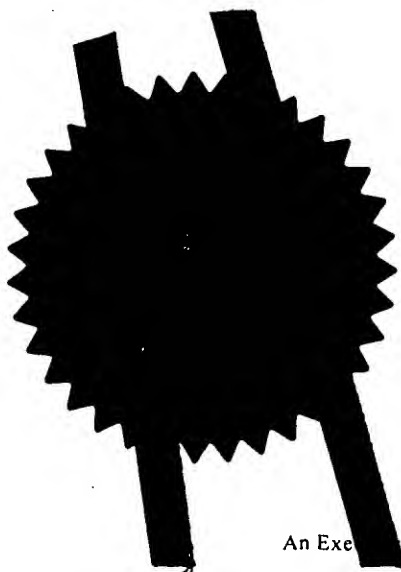
Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

CERTIFIED COPY
PRIORITY DOCUMENT

Signed

Dated

24 FEB 2002



An Executive Agency of the Department of Trade and Industry

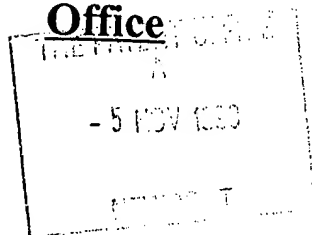
**The
Patent
Office**

7/77

Patents Act 1977

Rule 15

Statement of inventorship and of
right to grant of a patent



The Patent Office

Concept House
Cardiff Road
Newport
South Wales NP9 1RH

1. Your reference UK999109

2. Patent application number
(if you know it) **9926134.9** 05 NOV 1999

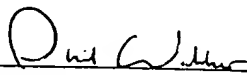
3. Full name of the or of each applicant INTERNATIONAL BUSINESS MACHINES CORPORATION

4. Title of invention INTERACTIVE VOICE RESPONSE SYSTEM

5. State how the applicant(s) derived the right from the inventor(s) to be granted a patent By employment and by agreement

6. How many, if any, additional Patents Forms 7/77 are attached to this form?

7. I/We believe that the person(s) named over the page (and on any extra copies of this form) is/are the inventor(s) of the invention which the above patent application relates to.


Signature

4 November 1999
Date

8. Name and daytime telephone number of person to contact in the United Kingdom P Waldner
01962 816057

Enter the full names, addresses and postcodes of the inventors in the boxes and underline the surnames

J B PICKERING

(Resident of the UK)

c/o IBM United Kingdom Limited
Intellectual Property Law
Hursley Park
Winchester
Hampshire
SO21 2JN
UK

Patents ADP number (if known)

07711997002

Patents ADP number (if known)

If there are more than three inventors, please write their names and addresses on the back of another Patents Form 7/77 and attach it to this form

Patents ADP number (if known)

REMINDER

Have you signed the form?

The Patent Office

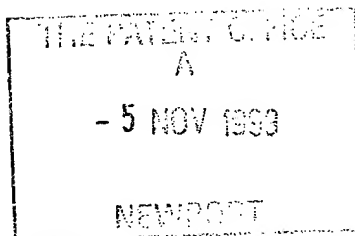
05NOV99 E489495-1 D00611
P01/7700 0.00-9926134.9

1/77

Patents Act 1977

Rule 16

Request for grant of a patent



The Patent Office

Concept House
Cardiff Road
Newport
South Wales NP9 1RH

| | | | | |
|----|--|---|--|---|
| 1. | Your reference | UK999109 | | |
| 2. | Patent application number (The Patent Office will fill in this part) | 9926134.9 | | 05 NOV 1999 |
| 3. | Full name, address and postcode of the or of each applicant (<i>underline all surnames</i>) | INTERNATIONAL BUSINESS MACHINES CORPORATION Armonk New York 10504 United States of America | | |
| | Patents ADP number (<i>if you know it</i>) | 07534092001 | | |
| | If the applicant is a corporate body, give the country/state of its incorporation | State of New York United States of America | | |
| 4. | Title of the invention | INTERACTIVE VOICE RESPONSE SYSTEM | | |
| 5. | Name of your agent (<i>if you have one</i>) | P Waldner | | |
| | "Address for Service" in the United Kingdom to which all correspondence should be sent (<i>including the postcode</i>) | IBM United Kingdom Limited Intellectual Property Department Hursley Park Winchester Hampshire SO21 2JN | | |
| | Patents ADP number (<i>if you know it</i>) | 07104417001 | | |
| 6. | If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (<i>if you know it</i>) the or each application number | Country | Priority App No (<i>if you know it</i>) | Date of filing (<i>day/month/year</i>) |
| 7. | If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application | No of earlier application | | Date of filing (<i>day/month/year</i>) |

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.)

Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description

12

Claim(s)

2

Abstract

1

Drawing(s)

6

10. If you are also filing any of the following, state how many against each item.

Priority documents

0

Translations of priority documents

0

Statement of inventorship and right to grant of a patent (Patents Form 7/77)

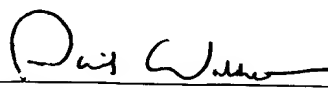
2

Request for preliminary examination and search (Patents Form 9/77)

Request for substantive examination (Patents Form 10/77)

Any other documents (please specify)

11. I/We request the grant of a patent on the basis of this application


Signature

4 November 1999
Date

12. Name and daytime telephone number of person to contact in the United Kingdom

P Waldner
01962 816057

INTERACTIVE VOICE RESPONSE SYSTEM

FIELD OF INVENTION

5 This invention relates to an interactive voice response system using speech recognition as input. In particular it relates to a method of dialogue management in the interaction using the statistical distribution of the type of words used by the caller interacting with the system. The management system makes decisions regarding the prompt selection and call flow of the dialogue.

BACKGROUND OF INVENTION

15 An interactive voice response (IVR) system is a computer system integrated with a telephone system that allows a caller to dial into the computer system over a telephone line and access a service running on the computer. The caller may then interact with and receive voice information from the service. Typically the interactive service has a range of services for the caller to choose from and presents options at a prompt menu expecting the caller to select one. After the service option has been chosen further information is required from the caller and input to the IVR. The service gathers relevant information, processes the information to get a result and from the result creates a prompt for delivery to the caller.

25 The interaction between the users and the system comprises various voice prompts output by the system and responses thereto input, via the telephone keypad, by the user. Voice response systems are used by service providers, such as banks, to automate fully or partially telephone call answering or responding to queries. Typically a voice response system provides the capability to play voice prompts comprising recorded voice segments or speech synthesised from text and to receive responses thereto. The prompts are generally played as part of a voice menu invoked by the call flow logic. A state table can access and play a voice segment or synthesise speech from given text. The prompts are usually part of a voice application which is designed to, for example, allow a customer to query information associated with their various banks accounts.

40 As the users of such system may not be familiar with the use thereof, it is necessary to ensure that the instructions or voice prompts are sufficiently comprehensive to allow a novice user to successfully interact with the system. However, the more competent the IVR user the more they begin to anticipate the various voice prompts and it becomes increasingly tedious for them to have to listen to such comprehensive instructions when more succinct instructions would suffice. "Expert" or

fastpath methods are often provided, usually on explicit user selection. These allow the caller to enter multiple pieces of information at one time, to hear shorter and more succinct prompts.

5 European patent publication 0697780 discloses a system for varying the voice menus and segments presented to the user of a voice response system according to the competence of the user. The response time of a user to voice prompts is measured and an average response time is
10 determined. It is assumed that the lower the average response time, the greater the competence of the user. The average response time is used as an index to a table of ranges of response times. Each range has
15 respective voice segments associated therewith. The voice segments comprise oral instructions or queries for the user and vary according to the anticipated competence of the user. If the average response time
20 changes such that the voice segments indexed are different to the current voice segments then a data base containing information relating to user competence is updated to reflect such a change. Accordingly, when the user next interacts with the voice response system a new set of voice segments more appropriate to the user's competence will be played.

 Using response times as a gauge of caller competence is only a first approximation and can be incorrect. The above publication concentrates on dual tone multi-frequency (DTMF) input to the IVR which is accurate but limited to a sometimes tedious closed menu structure and
25 set sequences. A more flexible but less accurate approach to caller interaction uses speech converted into text as input to a service. For instance instead of presenting the caller with an audible menu the caller can be asked a more open question as to the nature of his business. An automatic speech recognition component (ASR) translates the speech into
30 text and the IVR interprets the text in the light of the services offered. Response times have only limited effectiveness as an estimate of the competence of the caller in such circumstances. It is an object of at least the embodiment of the invention to provide an improved method of estimating the competence of a user for a speech recognition IVR system.

35 With the advent of advanced language processing techniques, such as Natural Language Understanding and Dialogue Management, the potential for both the "expert", and less-experienced users to benefit from fastpaths and task switching are increasing. However, there are two crucial
40 factors which cannot be catered for easily. Firstly, it is necessary to introduce an explicit method such as menu selection, or even Caller Line Identification, to switch between expert and novice versions of a service. Secondly, this selection is made on a service-wide basis, and does not change either without redialling or returning to a point in the
45 service where the selection may be made. Caller Line Identification (CLID) (or Automatic Number Identification (ANI) in the US) can be used

to retrieve caller records in which is stored the caller's preference of export or notice prompt. When given the choice many callers will wrongly identify themselves as "expert" and discover that the service is not responding well because the service has changed or because of environmental factors. There is a common assumption that experts will automatically use or want to use barge-in, and the novice not. But again, for environmental reasons, each group is better served by flexibility.

The selection of expert or novice prompts, however, is not a generically applicable distinction, which the caller themselves can necessarily judge. In some circumstances (background or channel noise), the caller would be better served as a novice. Competence may change within the same application which increases the problem of how to define 'expert'. Expertise might be regarded as some level of competence in achieving a given task in the most efficient manner.

SUMMARY OF INVENTION

In one aspect of the present invention there is provided a method of managing dialogue in an interactive voice response system comprising the steps of: receiving a voice signal from a caller to the IVR; converting the voice signal to text; estimating a caller type based on the number of words of a particular type within the text; using this factor to make dialogue decisions.

Therefore when a caller instigates an interaction with the voice system and voices a response, the IVR takes this response and undertakes a statistical analysis of words of a certain type. The analysis is used as a basis for making decisions on the flow and structure of the dialogue with the caller.

In the embodiment a decision on which prompt to present next to the caller is made depending on the factor and in particular a decision whether to use expert or novice prompts is made depending on whether the factor is above or below a threshold value. The expert may be viewed not necessarily as someone who has used a service many times and remembers what the answer to the next prompts will be, but rather a person who effectively gets what they want on a single pass from the automatic service: the expert is the user who, regardless of background or channel interference, most effectively achieves what they set out to do. More advantageously a decision whether to use expert, intermediate, or novice prompts is made depending on whether the factor falls below, inside or above a threshold range.

5 An important and useful factor is a ratio of words relevant to the
context of the dialogue which gives a measure of the communicative
efficiency of the incoming speech. For instance an expert caller would
know and use more relevant words in a dialogue than a novice who had yet
to learn what words were the most relevant. It is a factor derived from
the ratio of lexically marked and unmarked items and gives a first
approximation of the competence of the caller. Another factor is the
number of words per second spoken by the caller which gives an indication
10 of the "regularity" of delivery (the prosodic fluency). For instance an
expert caller would be confident and regular in delivery of speech
whereas a novice would be somewhat hesitant. The word rate factor may be
used alone or most advantageously is used together with the context ratio
to give a further approximation as to the competence of the caller.
15 Ideally the context ratio and the word rate factors are weighted and
combined to give an overall factor having a threshold value or range to
decide the level of competence of the caller.

20 Another factor is a ratio of the words relevant to the particular
task or service which is part of the dialogue. This gives an indication
of the task effectiveness of the incoming speech. The task words are a
subset of the context words which are appropriate to the current task. It
is important therefore that the IVR is aware of an individual task within
the dialogue and what group of task words corresponds to the tasks. The
IVR application services correspond with the groups of task words.
25 Advantageously the context ratio, word rate factor and service ratio are
weighted and combined to give an overall factor having a corresponding
threshold value or range.

30 Another factor is a ratio of extra linguistic items in the text
such as 'er', 'um' and unrecognized items. The context ratio, word rate
factor, service ratio and extra linguistic items are weighted and
combined to give an overall factor. Extra linguistic terms are
irrelevant to LEXICAL meaning.

35 The context ratio, word rate factor, service ratio and extra
linguistic items are weighted and combined to give an overall factor.

40 Another factor is a ratio of stylistic items in the text such as
'should', 'like' and 'gimme'.

45 Combining all the factors together would indicate whether or not
the caller is an "expert" and therefore whether or not the application
structure and the outgoing prompts could be modified to ensure caller
satisfaction and the efficient running of the service tailored to the
caller's needs. Further, a running average is maintained throughout a
call: moment-to-moment excursions from this could indicate difficulties

which should be addressed; further, variance would be used to indicate chance expert interactions which might be ignored in the automatic decision as to whether or not the caller is an expert and therefore can be allowed to be left with little "intervention" from more excursive prompts.

BRIEF DESCRIPTION OF DRAWINGS

In order to promote a fuller understanding of this and other aspects of the present invention, an embodiment will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1 is a schematic representation of the IVR system of the present embodiment;

Figure 2 is an example prompt voice database

Figure 3 is an example prompt script database;

Figure 4 is a schematic of a lexical analyzer;

Figure 5 is the method of the present embodiment; and

Figure 6 is an example of a dictionary of the present embodiment.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENT

Referring to Figure 1 there is shown a schematic representation of the embodiment. An interactive voice response 10 system including standard IVR capabilities is connected to a LAN 12 network to access server-based automated services to perform specialized tasks. A business server 14 is connected to the LAN 12 to perform the business function part of the service. An automatic speech recognition (ASR) server 16 is a speech to text engine and is connected to the LAN 12. A natural language unit (NLU) 18 interprets the converted text to extract relevant words as data items. A dialogue manager (DM) 20 connected to the LAN 12 provides prompts to the IVR 10. A profile server 22 provides caller profiles to the DM 20 for construction of personalized prompts. The IVR 10 also comprises a lexical analyzer 50 that estimates the communication efficiency of the caller interaction and uses this factor to chose a competence of the caller. Based on this the IVR selects a set of prompts to play to the caller.

The IVR 10 is connected to a telephone or Internet network 24 which in this embodiment is a land line but which in other embodiments is a

5 wireless telephony network or an IP telephony connection. A caller 26
accesses the IVR 10 via the telephone or Internet network 24. The IVR
provides voice messaging and audio text, whereby the caller interacts
with the automated service via option menus to decide what information is
required and what service is needed. The IVR normally provides several
services but in this embodiment one is described for simplicity. The IVR
runs the call flow through a dialogue in conjunction with the DM 20 which
supplies the prompts. In this embodiment the IVR and DM 20 are described
as separate servers but are closely related functions and may be
10 implemented together in other embodiments. The IVR 10 provides the
interface to the back end services such as the business server 14. It
provides explicit or implicit caller identification by calling line
identification (CLID) or through explicit sub dialogue with the caller.
When a caller requests a service the IVR 10 opens a service record 28
15 containing service record fields 30 that require data. The IVR collects
pieces of information from the caller or from caller account (available
from the business server 14) and fills the service record fields 30.
Once complete the service record 30 may be sent to the business server 14
to process the record and deliver an answer. An example of such a voice
20 response system is the IBM Corepoint Voice Response for AIX product.

The business server 14 executes a business application that may
process a service record containing information in fields. This is an
end step of the embodiment and requires that all the information for a
25 particular service has been acquired before it carries out its function.
The business server is arbitrary depending of the application of the IVR
and the embodiment is focused on collecting service information rather
than processing the information.

30 The ASR 16 server provides high function speaker-independent
recognition capabilities including the following: dynamic grammar (or
similar) loading and generation; dynamic application dictionary (or
equivalent) loading; optionally, and if appropriate, modification of any
stochastic language models; grammar and/or dictionary sub setting; large
35 vocabulary; continuous speech input; and speaker independence. The ASR
functionality can be installed and running in the client device, in which
case there needs to be a data exchange protocol between the client device
and the "server", such that the client can upload relevant information to
the server for tracking purposes. The ASR also provides a length of
40 sentence parameter and a hit-rate or confidence score parameter on the
response.

The natural language understanding unit (NLU) 18 extracts the data
information from a text response by tagging the meaning of the text.
45 More than one piece of data may be identified in the response and it is

the function of the NLU to isolate and extract important data and pass them on to the dialogue manager (DM) 20.

5 The dialogue manager (DM) 20 comprises the service record 28 and associated service record fields 30; a prompt generator 32 and a prompt database 34 (see Figure 2). The DM 20 also has the facility to track: the words (lexical items) used by the caller; any associated synonyms; the path taken through the call flow by the caller; and the grammatical (syntactic) structure of caller input.

10 The prompt generator 32 takes a prompt script and generates a playable prompt in the form of a single voice data packet for the IVR. A prompt script is represented by the mnemonic code $P(x,y)$ where 'x,y' represents the contents of the playable script and identifies one or more prompt parts in a prompt database array. For instance, $P(N,p1)$ is a playable prompt representing the novice prompt p1. The prompt generator will acquire the voice data packet for novice prompt p1 and construct a playable prompt. When the contents of the script are a combination of prompt parts the individual parts will be combined to make a playable prompt. For example, $P(N,p1+p2)$ is a playable script representing a combination of the novice prompt parts p1 and p2. The prompt generator, on receiving this script, will fetch the individual prompt parts and combine them as required so that they play as one. For example, novice prompt part p1 "What is the source account?" and novice prompt part p2 "What is the destination account?" are combined with an "and" voice data packet to render the combined prompt of "What is the source account and what is the destination account?". In another embodiment prerecorded combination prompts could totally or partially replace on the fly generation of combination prompts from individual prompts.

30 The profile server 22 contains the caller profiles including the prompt scripts (see Figure 3). Each profile contains key indicators, such as the IVR retry rate, IVR time out rate, vocabulary subsection, grammar subsection, and expected call flow route. Each profile is maintained and made available to the IVR service. The profile server 22 may also hold other caller or customer databases such as an account database that can be relationally linked to the caller prompt profiles.

40 The prompt database 34 comprises the voice data prompts used by the IVR to interact with the caller (see Figure 2). A prompt is a voice recording relating to a specific IVR request and a service field in a service record has a corresponding prompt that is played to the user when the data item for the service field is required. The prompts are stored and indexed for retrieval by the prompt generator when it processes a prompt script. The prompts are categorized by caller competence so that for each type of request there are corresponding expert, intermediate and

novice types of prompt. For instance, if the prompt generator needs to play the first prompt in order to ascertain which is the caller's source account from which money is to be transferred it could select from the expert prompt of "Source?", the intermediate prompt of "What is the source account?" and the novice prompt of "What is the source account that you wish to transfer money from?". The criteria for whether the caller is an expert, intermediate or novice can be stored in the caller profile. In this embodiment there are four prompts but in other embodiments there can be many services with varying numbers of prompts. In this embodiment each prompt corresponds to a single request, for multiple requests the prompts need be combined in the prompt generator. In other embodiments single prerecorded prompts may represent multiple requests without the need to combine them.

The prompt script database 36 comprises a script associated with a particular caller and a service record for a number of different types of callers and also individual callers (see Figure 3). A new or unidentified caller is given the novice default script of $P(N,p1)$, $P(N,p2)$, $P(N,p3)$ and $P(N,p4)$ without any combined or reordered prompts. Each playable prompt will be based on the original sequence of individual prompt parts. An example of a caller using intermediate multiple prompts is $P(I,p1+p2)$, $P(I,p3+p4)$ whereby the prompt generator will combine the intermediate $p1$ and $p2$ prompt parts as a single playable prompt voice data packet and then combine the intermediate $p3$ and $p4$ prompt parts as a single playable prompt voice data packet. An example of a caller using multiple reordered prompts is $P(E,p2 + p1)$, $P(E,p4 + p3)$ whereby the prompt generator will combine the expert prompt parts in a different order than previously mentioned. A caller matched by calling line identification number (CLID=01962 815000) has mixed combined and individual prompts scripts $P(p1 + p2)$, $P(p3)$, $P(p4)$. The embodiment allows such a combination to be unique to that caller. In this embodiment the caller will have the actual script associated with his caller id but in another embodiment there may be an indirect association using a pointer to a common type of script.

The lexical analyzer 50 (see Figure 4) comprises a search engine 52 for searching dictionary 54 against ASR text; a count engine 56 for counting the number of words in each lexical category; a ratio calculator 58 for determining the ratio of the lexical types and decision logic 60 for determining the competence of the caller.

Dictionary 54 is a lookup table with two fields: word and corresponding category code. The category code determines in which lexical categories the word belongs using binary indicators in sixteen or more bit number. The lexical categories are: content words; task words; function words; extra linguistic words; paralinguistic words and

stylistic words. Content words are words which carry lexical meaning ("account", "transfer", and so forth, particular to the business application, and words such as "tree", "telephone", etc.). Task words are those specific to a particular service, and a subset of all content words. For instance, in a banking application: 'bank balance'; 'transfer'; 'statement'; 'account'; 'operator'; 'overdraft'; 'date'; 'percent'; 'pounds'; 'dollars'. In a banking application with several services there will be a corresponding number of task word subsets of content words for making a transfer between accounts; ordering a statement; or checking a balance etc. Function words are words which are not content words but add to the grammatical correctness of the phrase, for example, 'of', 'the', 'and', 'a', 'when', 'how', 'is', 'was' and other auxiliary words, articles and conjunctions. Phatic words add no specific lexical or grammatical information, but affect the perception of the sentence (such as "oh"). Stylistic words include words like "should" and "gimme". In this embodiment five distinct categories and one task category are used but more distinct and more task categories may be used with a corresponding increase in the size of the category code. Figure 6 is an example dictionary with word types indicated.

Search engine 52 is for extracting a word from the ASR text and querying the dictionary for the category code which it passes on to the count engine.

Count engine 56 is for counting the number of words in the ASR text belonging to each category. It comprises a counter for each distinct lexical category and each separate task category: a content word counter; a task counter; a function word counter; an extra linguistic word counter; a paralinguistic word counter and a stylistic word counter. It also comprises a total word counter. Each bit of the category code is checked and the corresponding counter is incremented if set. After the ASR text has been analyzed the counters contain the final result.

Ratio calculator 58 takes the results of the count engine 56 and works out a structure ratio; a service ratio; an extra linguistic ratio; a paralinguistic ratio and a style ratio. The structure ratio is the ratio of content words to function words. The service ratio is the number of phatic words in the text divided by the task words for the service that the IVR is using at that time. The ratio calculator queries the IVR as to the particular service and chooses the correct task counter accordingly. The extra linguistic ratio is the number of extra linguistic words in the ASR text to the total number. The paralinguistic ratio is the number of paralinguistic words divided by the total number of words. The stylistic ratio is the number of stylistic words divided by the number of the total number of words.

Decision logic 60 takes its input from the ratio calculator 58 to decide the level of competence of the caller. For example, in one embodiment the decision logic uses the structure ratio only; if the ratio is 20% or below then the caller is deemed unfocused in his use of the correct type of words (content words), and is given novice status; if the ratio is above 20% then the caller is deemed competent in the use of content words and given expert status. In an enhancement to the first embodiment an intermediate level of competence is defined for above 15% and below 31% with novice and expert level either side of this range. In a further enhancement the decision logic uses the service ratio only; if the ratio is 70% or below the caller is deemed a novice; if above 70% then the caller is deemed an expert. An intermediate level is defined between 65 and 76% exclusive with novice and expert levels either side of this ratio. Similar embodiments may make use of the extra linguistic ratio; paralinguistic ratio and stylistic ratio on their own to decide between expert, intermediate and novice. A further more useful embodiment uses a ratio matrix of the structure ratio and service ratio and multiplies them by a weighting matrix to get an overall factor for which there is a threshold novice / expert value and a threshold intermediate range. Embodiments making further approximations include one or more of the extra linguistic ratio, paralinguistic ratio and stylistic ratio in the ratio matrix with an appropriate weighting matrix and threshold value. Note: since words within the ASR result string will typically fall into several categories (e.g., "account" would be both content and task word, the various ratios for a given input sentence cannot simply be added with an expectation of producing 100% for the complete expert. The example below will illustrate.)

The process steps of the lexical analyzer are described with respect to Figure 5. After dialing into an IVR 10 a caller speaks a response which is recorded and processed by the ASR to get a text string (start 5.1). The text string is retrieved (5.2) by the lexical analyzer and a first words or words identified (5.3) by segmentation of the text string by identifying groups of the characters which are separated by spaces. A query is sent (5.4) to the dictionary 54 to locate the first word. A category code is extracted (5.5) from the dictionary and returned to the lexical analyzer 50. The code is used by the counter engine to increment (5.6) the word counters for the array of word types. The next word and subsequent words in the text string are dealt with in a similar manner (5.7) by looping back to the extract word step using the next and subsequent words and searching (5.4) the dictionary as before. Once all the words have been queried using the database the ratios are provided (5.8) by the ratio calculator 58 by dividing certain counts by corresponding counts. The decision logic 60 takes input from the calculator of the structure ratio, the particular service ratio corresponding to the IVR prompt, the extra linguistic ratio, the

paralinguistic ratio and the stylistic ratio to estimate a competence value and select (5.9) a competence level. The result is fed back to the IVR so that the next prompt can be personalized to the caller.

5 Consider the following example, based on the words and word types of Figure 6, in which a caller wishes to find out the balance on their account. They may use the following constructs:

10 I. *Erm I'd like to check my account balance, please*

The "...." (a noticeable pause) is discounted for simplicity. This construct then includes ten "words", as follows:

| | | | | |
|------------------|-------------------------------|------|-----|-------------------|
| Extra linguistic | ERM | 1/10 | 10% | --> NOVICE |
| Function | I, would, to my | 4/10 | 40% | --> NOVICE/INTERM |
| 15 Phatic | I, (woul)d, like, to, please | 5/10 | 50% | --> NOVICE |
| Content | like, check, account, balance | 4/10 | 40% | --> INTERMEDIATE |
| Task | account, balance | 2/10 | 20% | --> NOVICE |
| Task : Content | | | 50% | --> NOVICE/INTERM |

20 This gives 3 straight NOVICE decisions; 2 mixed NOVICE or INTERMEDIATE; and 1 INTERMEDIATE. The overall decision is NOVICE.

II. *Give me my balance, please*

| | | | | |
|---------------------|---------------|-----|-----|-------------------|
| 25 Extra linguistic | -- | 0/5 | 0% | --> INTERM/EXPERT |
| Function | me, my | 2/5 | 40% | --> NOVICE/INTERM |
| Phatic | please | 1/5 | 20% | --> INTERMEDIATE |
| Content | give, balance | 2/5 | 40% | --> INTERMEDIATE |
| Task | balance | 1/5 | 20% | --> NOVICE |
| 30 Task : Content | | | 50% | --> NOVICE/INTERM |

This gives 1 NOVICE; 2 mixed NOVICE/INTERMEDIATE; 2 INTERMEDIATE; and 1 mixed EXPERT/INTERMEDIATE. The overall decision is INTERMEDIATE.

35 III. *Balance enquiry: account number 1 2 3 4 5 6*

| | | | |
|------------------|-------------|------|------------|
| Extra linguistic | - | 0% | --> EXPERT |
| Function | - | 0% | --> EXPERT |
| Phatic | - | 0% | --> EXPERT |
| Content | <all> 10/10 | 100% | --> EXPERT |
| 40 Task | <all> 10/10 | 100% | --> EXPERT |
| Task : Content | | 100% | --> EXPERT |

This gives 6 EXPERT and an overall decision of EXPERT.

5 In this example, for construct III, the caller has provided more
than expected (i.e., provided both the indicator of the service
("balance enquiry") and the appropriate data (the account number)). It is
clear that a further weighting could be applied for such "over
10 efficiency" or explicit fastpath methodology. Further, no account is
taken in these examples of word rate (i.e., how quickly and fluently the
input is provided) which would allow further differentiation of the
overall decision on the type of prompt that should be played to the
caller. For example, the faster the rate of delivery in example II
above, without compromising the ASR result (i.e., without introducing
15 recognition failures), the more mixed NOVICE/INTERMEDIATE decisions would
shift to INTERMEDIATE.

20 An enhancement to the embodiment which does not use the lexical
analyzer calculates the overall word rate of the caller by acquiring a
response time from the IVR and counting the words from the ASR.

25 A further enhancement tracks the fundamental frequency of the
response and counts the number of discontinuities over the period of time
and checks for a final rise at the end of the signal. A preponderance of
these factors indicates an uncertain voice and a novice caller..

30 Although this approach is directed to telephony it is not so
restricted. Such a method could apply to any automated service, for
instance, text dialogue in an Internet application

Now that the invention has been described by way of a preferred
embodiment, various modifications and improvements will occur to those
person skilled in the art. Therefore it should be understood that the
35 preferred embodiment has been provided as an example and not as a
limitation.

CLAIMS

1. A method of managing dialogue in an interactive voice response system comprising the steps of:

receiving a voice signal from a caller to the IVR;

converting the voice signal to text;

estimating a caller type based on the number of words of a particular type within the text; and

using the caller type to make a dialogue decision.

2. A method as in claim 1 wherein the caller type is an indication of the competence level of the caller.

3. A method as in claim 1 wherein the caller type is based on a word ratio.

4. A method as in claim 1 further comprising locating a word type flag in a dictionary for each word in the text; incrementing a word type counter depending on the flag; and estimating a caller type using the word type counter.

5. A method as in claim 4 further comprising identifying more than one word type from the word type flag and incrementing more than one word type counter

6. A method as in claim 5 further comprising identifying from each bit of the word type flag whether a word is a certain type or not, each bit of the word type flag being associated with one word type.

7. A method as in claim 1 wherein a decision on which prompt to present next to the caller is made depending on the caller type.

8. A method as in claim 7 wherein a decision whether to use expert or novice prompts is made depending on whether the caller type is above or below a threshold value.

9. A method as in claim 8 wherein a decision whether to use expert, intermediate, or novice prompts is made depending on whether the caller type falls below, inside or above a threshold range.

10. A method as in claim 1 wherein one word type is made up of words relevant to the context of the dialogue.

11. A method as in claim 10 wherein another word type is made up of words relevant to the context of a task specific part of the dialogue.

5 12. A method as in claim 10 wherein the number of words per second spoken by the caller is used together with the context type to give a further approximation as to the competence of the caller.

10 13. A method as in claim 12 wherein the context ratio and the word rate factor are weighted and combined to give an overall factor having a threshold value or range to decide the level of competence of the caller.

14. A system for managing dialogue in an interactive voice response system comprising:

15 means for receiving a voice signal from a caller to the IVR;

an ASR for converting the voice signal to text;

20 means for estimating a caller type based on the number of words of a particular type within the text; and

means for using the caller type to make a dialogue decision.

25 15. A computer program product, stored on a computer-readable storage medium, for executing computer program instructions to carry out the steps of a method as claimed in claim 1.

ABSTRACT

INTERACTIVE VOICE RESPONSE SYSTEM

5 The present invention relates to an interactive voice response
system which statistically analyses word usage in speech recognition
results to select prompts for use in the interaction. The method
comprising converting a received voice signal to text and calculating
factors such as context and task word ratios and word rate; and using
10 these factors to make dialogue decisions as to the whether to use expert,
intermediate, or novice prompts depending on whether the factor falls
below, inside or above a threshold range

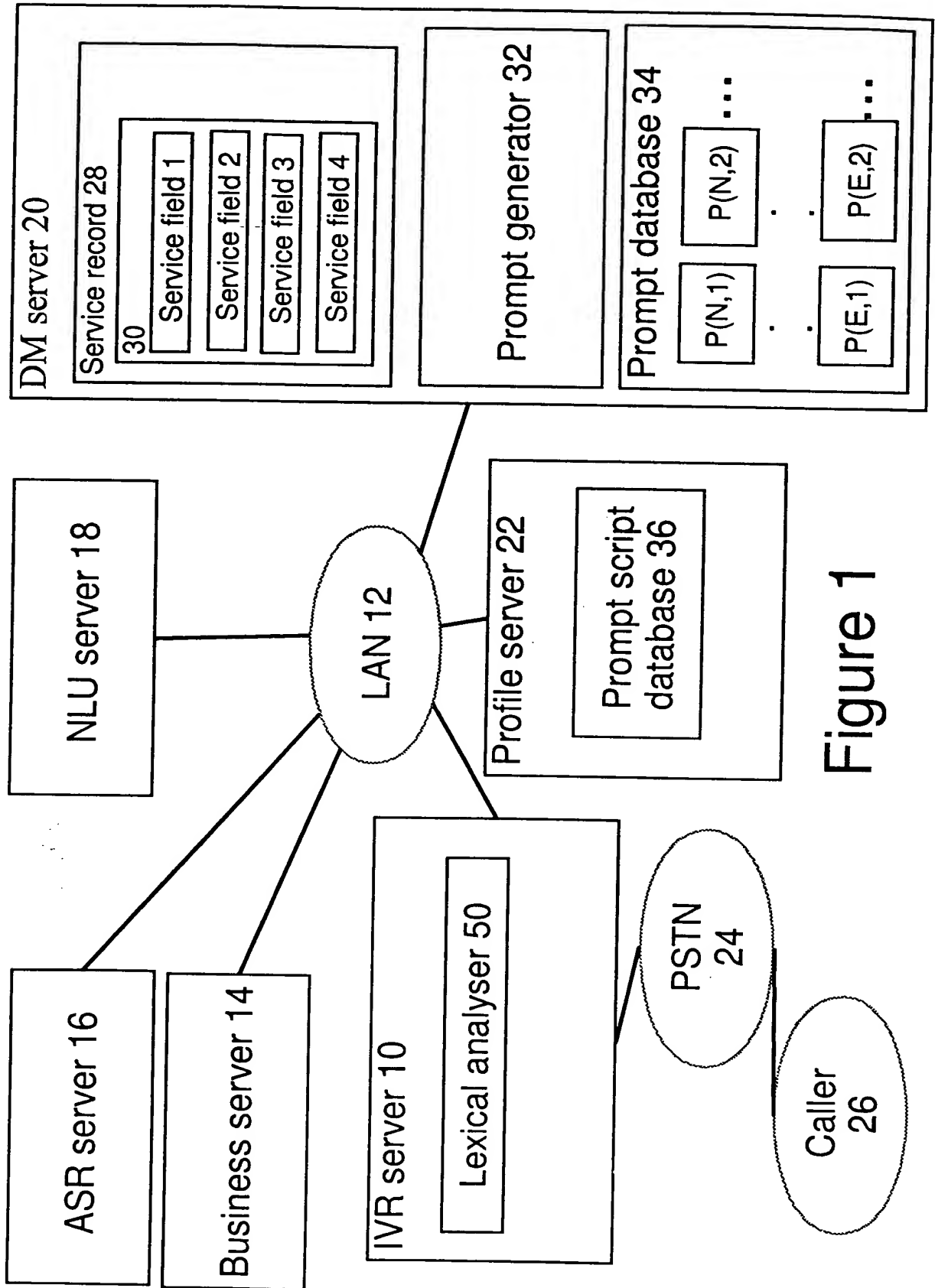


Figure 1

| Prompt database 34 | | | | |
|--------------------|--------------|----------------------------------|--|--|
| Prompt no. | Expert | Intermediate | Novice | |
| p1 | Source? | What is the source account? | What is the source account which you wish to transfer money from. | |
| p2 | Destination? | What is the destination account? | What is the destination account which you wish to transfer money to. | |
| p3 | Amount? | What is the transfer amount? | What is the amount of money which you wish to transfer? | |
| p4 | Timing? | When is the transfer date? | What is the date on which you wish the money to transfer to the destination account? | |

Figure 2

| | |
|----------------------------|--------------------------------------|
| Prompt script database 36 | |
| Caller no | Prompt scripts |
| Default caller | $P(N,p1); P(N,p2); P(N,p3); P(N,p4)$ |
| Multiple prompts | $P(l,p1 + p2), P(l,p3 + p4)$ |
| Multiple reordered prompts | $P(E,p2 + p1), P(E,p4 + p3)$ |
| caller CLI=01962 815000 | $P(N,p1 + p2), P(l,p3), P(E,p4)$ |

Figure 3

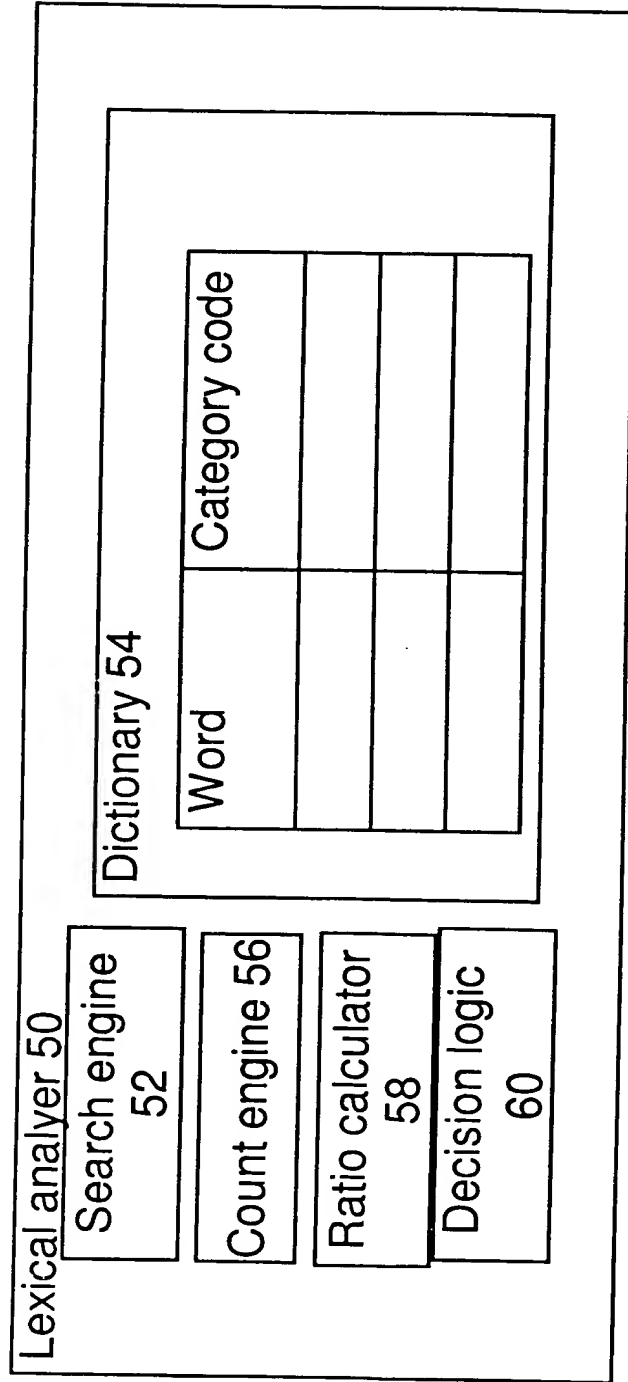


Figure 4

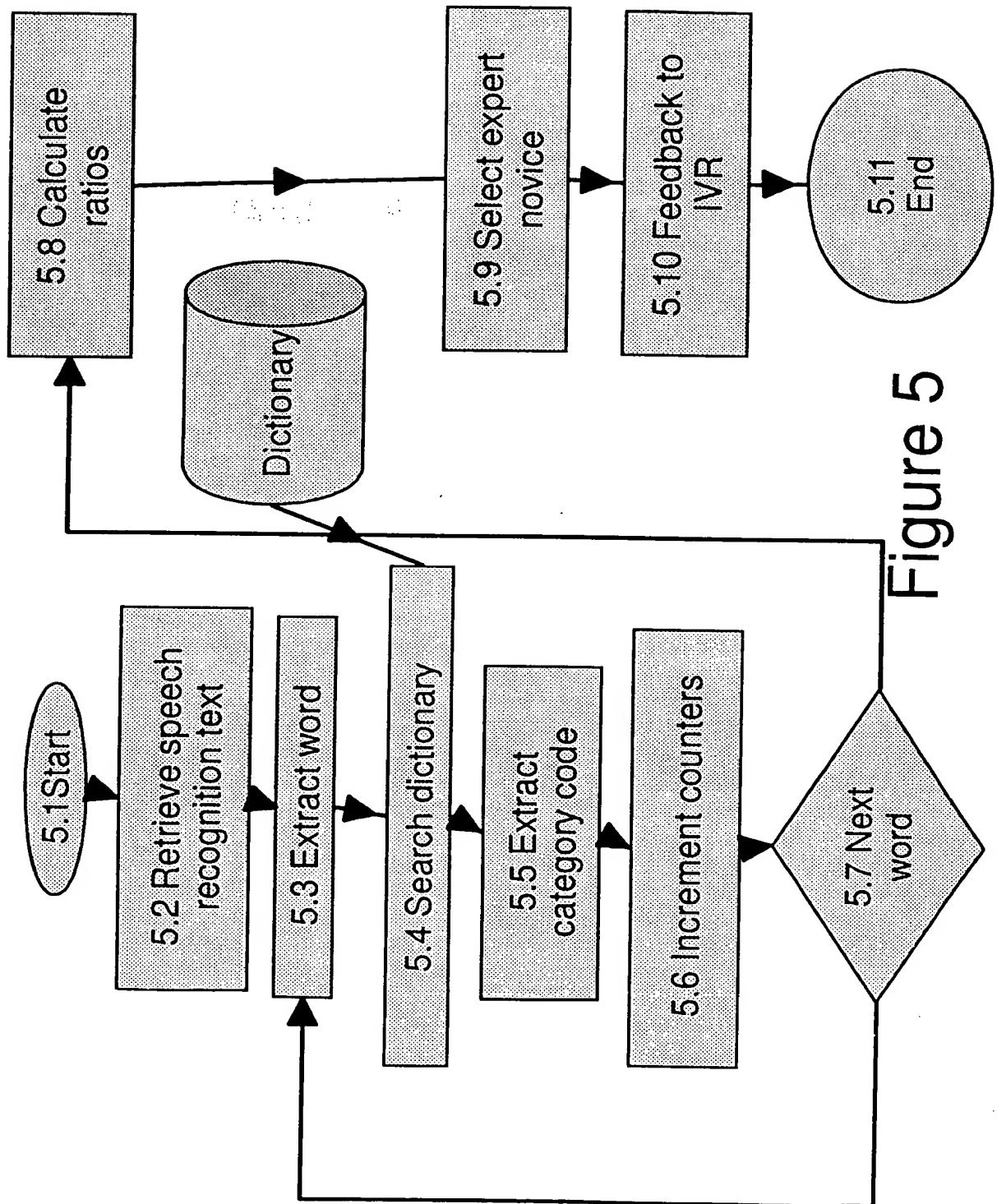


Figure 5

| Word | context | task | function | phatic | extra-linguistic | para-linguistic | stylistic |
|------------|---------|------|----------|--------|------------------|-----------------|-----------|
| I | | | | 1 | | | |
| want | | | 1 | 1 | | | |
| my | | | | 1 | | | |
| account | 1 | 1 | | | | | |
| balance | 1 | 1 | | | | | |
| which | | | 1 | 1 | | | |
| you | | | | 1 | | | |
| say | | | 1 | | | | |
| is | | | 1 | 1 | | | |
| over drawn | 1 | | | | | | |
| should | | | 1 | 1 | | | 1 |
| like | | | | 1 | | | 1 |
| bank | 1 | | | | | | |
| please | | | 1 | 1 | | | 1 |
| er | | | | | 1 | | |
| um | | | | | 1 | | |
| oh | | | | | | 1 | |

Figure 6